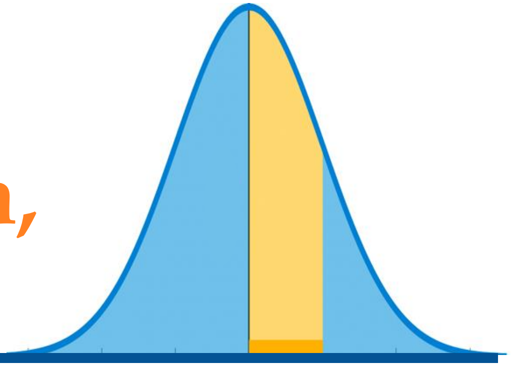
Efficient Statistics With Unknown Truncation,

Polynomial Time Algorithms, Beyond Gaussians



Jane H. Lee

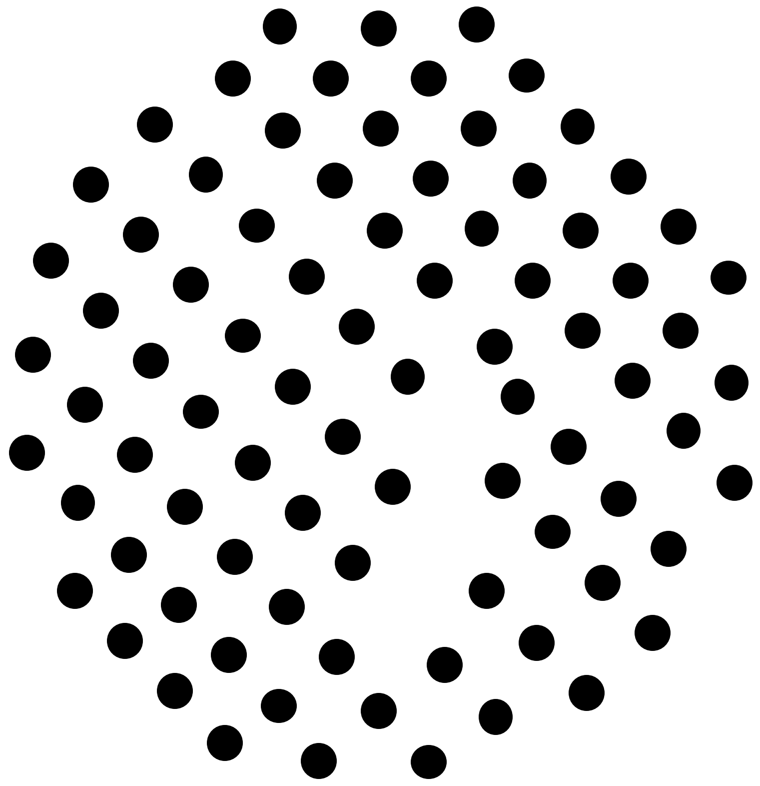
Anay Mehrotra

Manolis Zampetakis

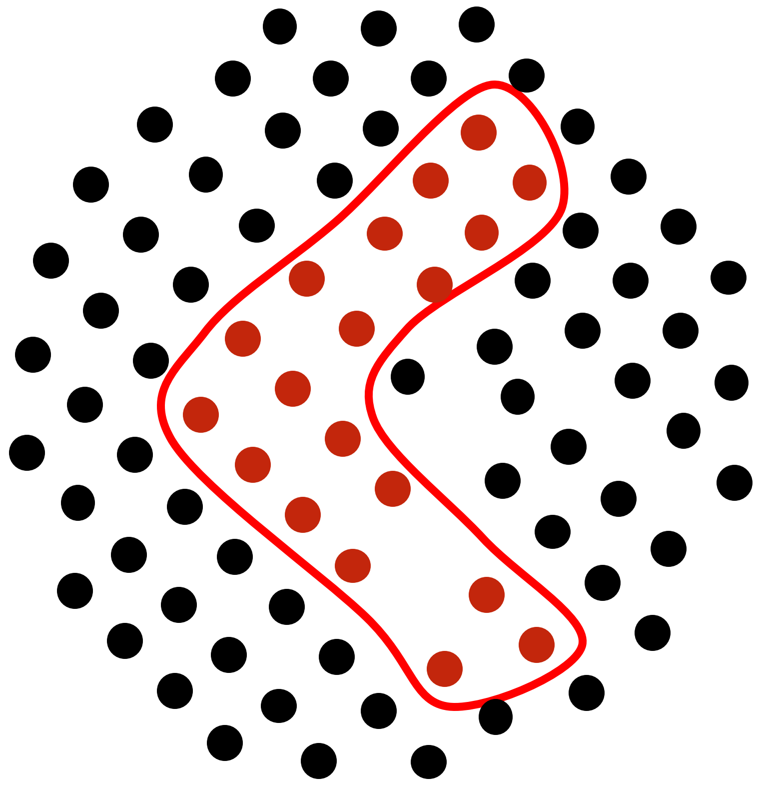
Yale

FOCS 2024

Estimation from Truncated Data

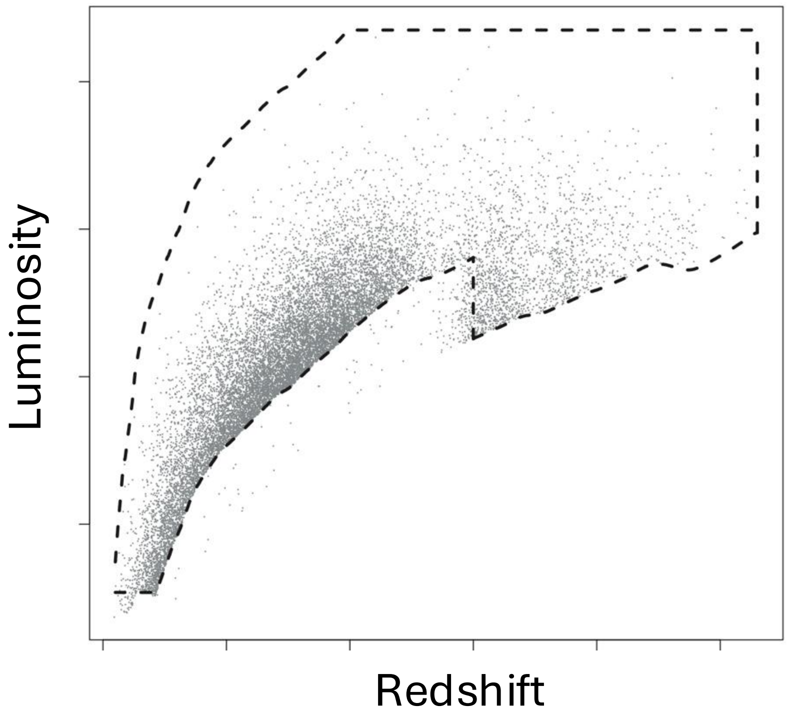


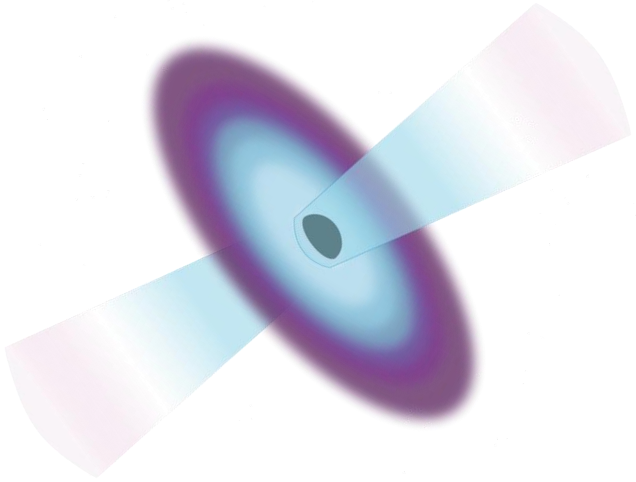
Estimation from Truncated Data



Truncation in Astronomy: Quasar Detection

Q: *Were older quasars brighter?*

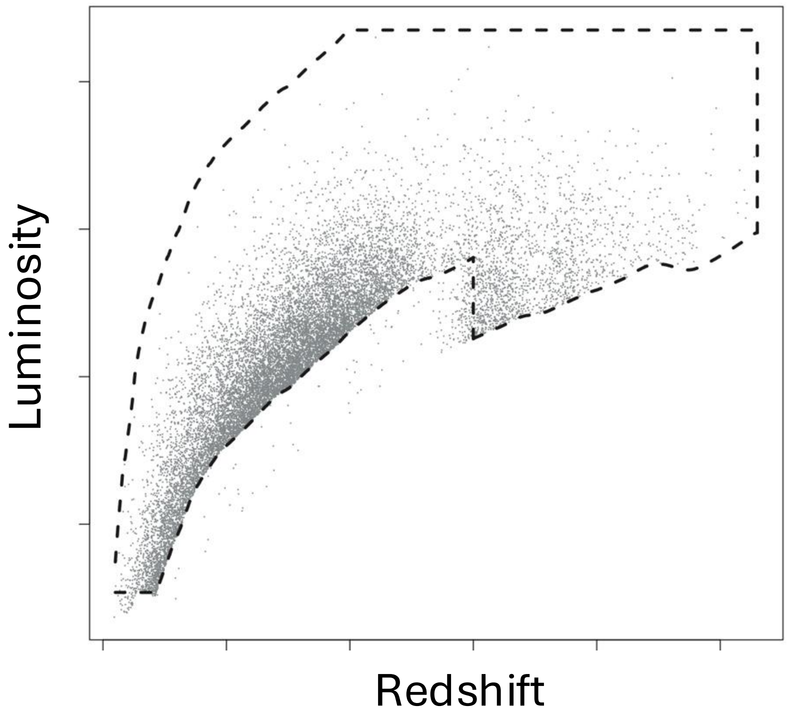


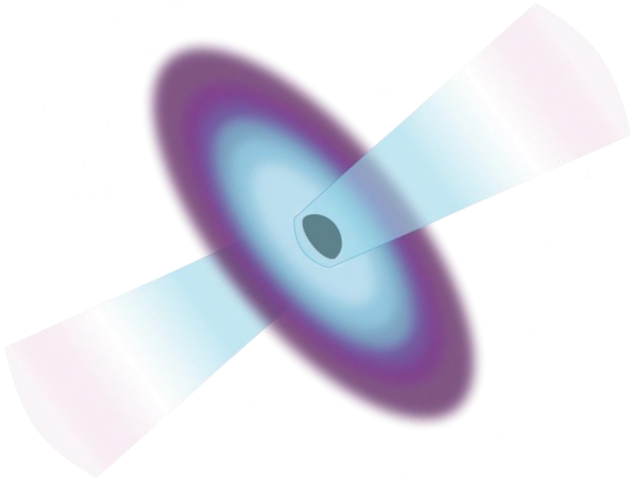


Truncation in Astronomy: Quasar Detection

Q: *Were older quasars brighter?*

Quasars are hard to detect! [Schafer, 2007]

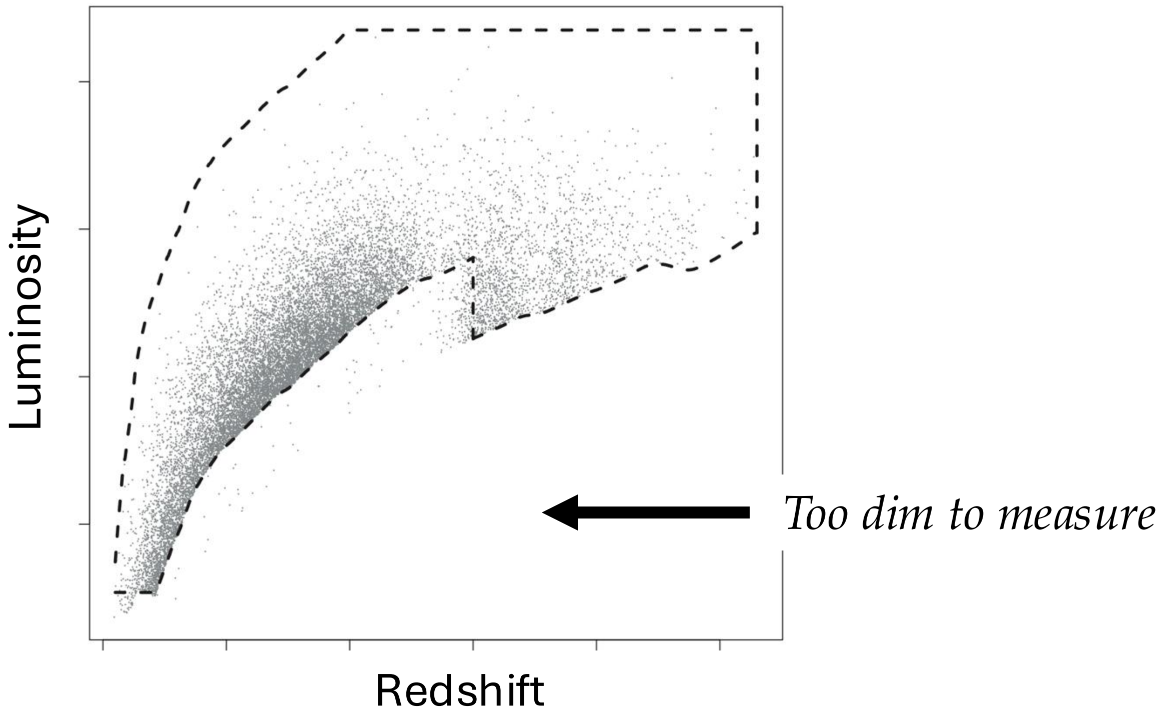


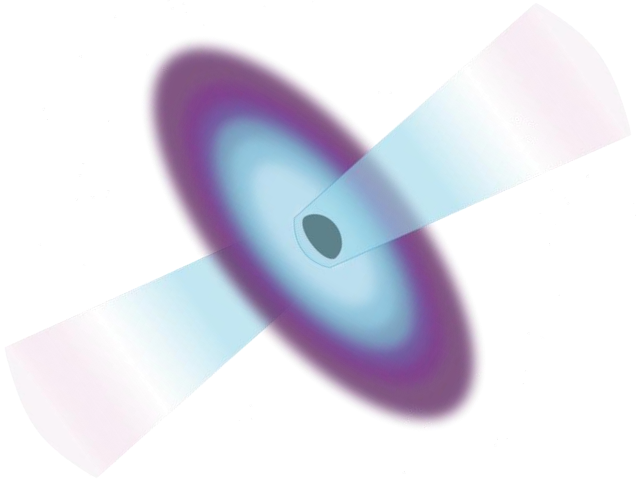


Truncation in Astronomy: Quasar Detection

Q: *Were older quasars brighter?*

Quasars are hard to detect! [Schafer, 2007]

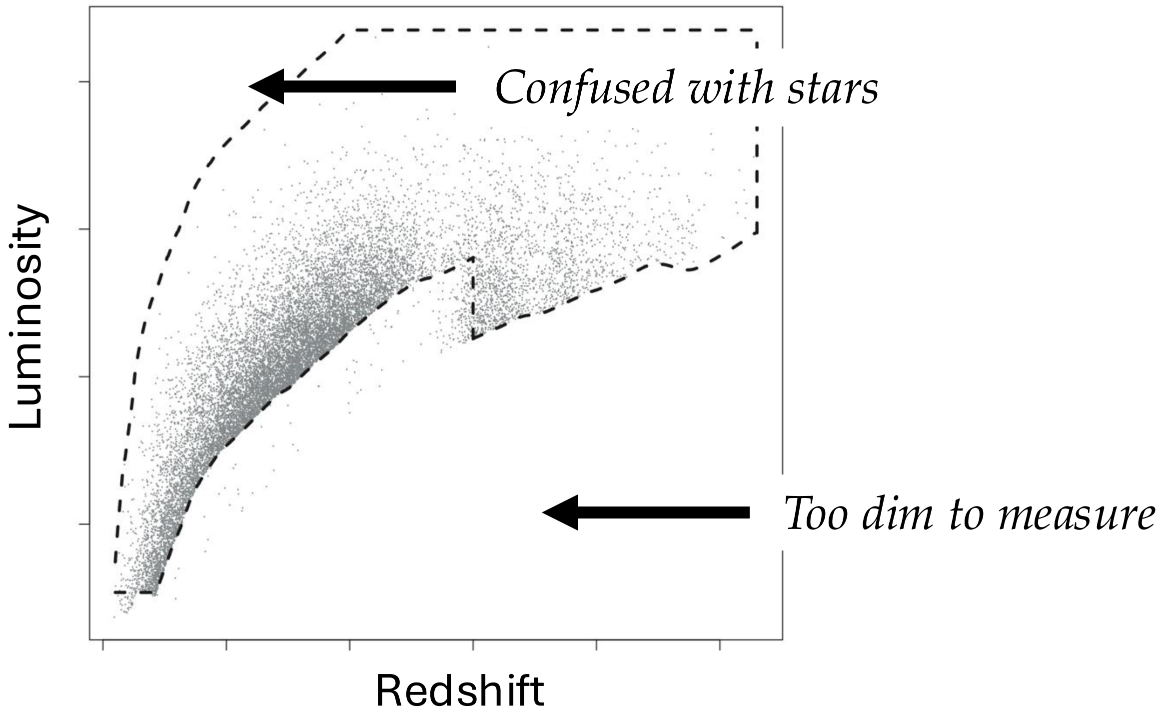


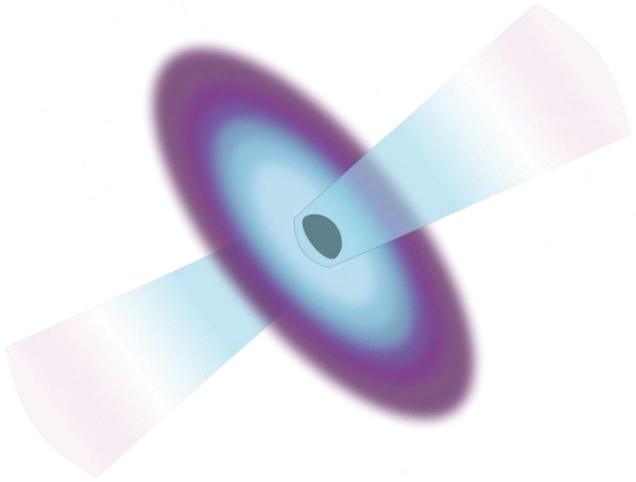


Truncation in Astronomy: Quasar Detection

Q: *Were older quasars brighter?*

Quasars are hard to detect! [Schafer, 2007]

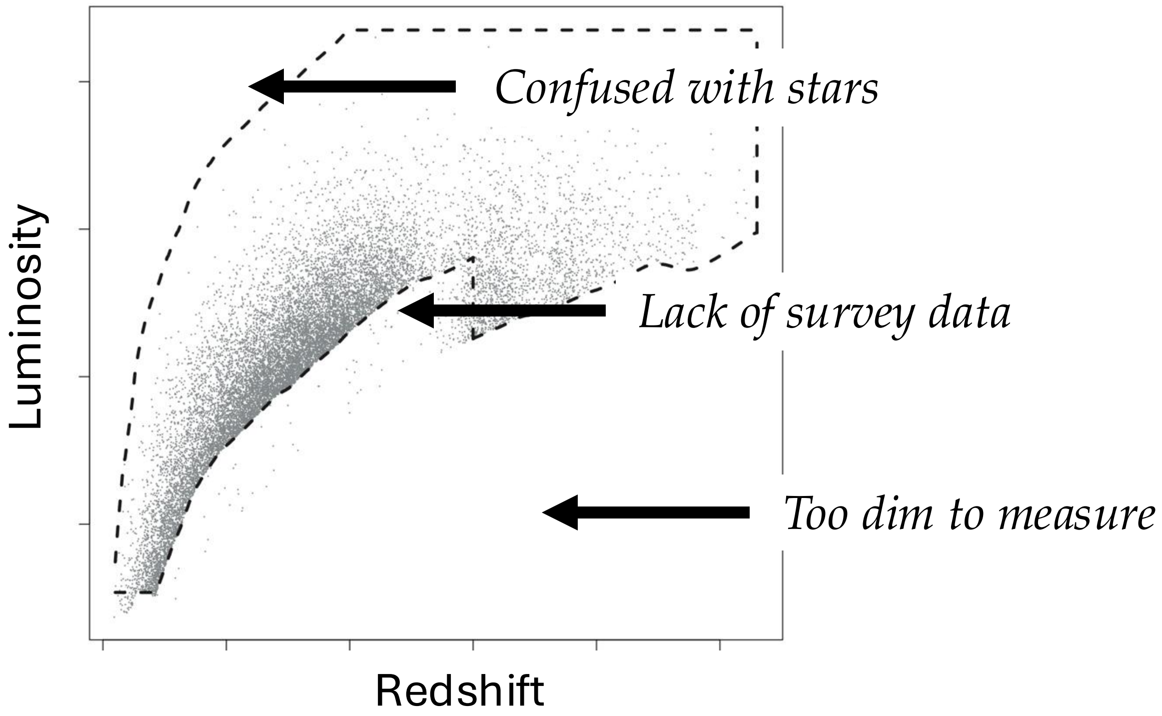


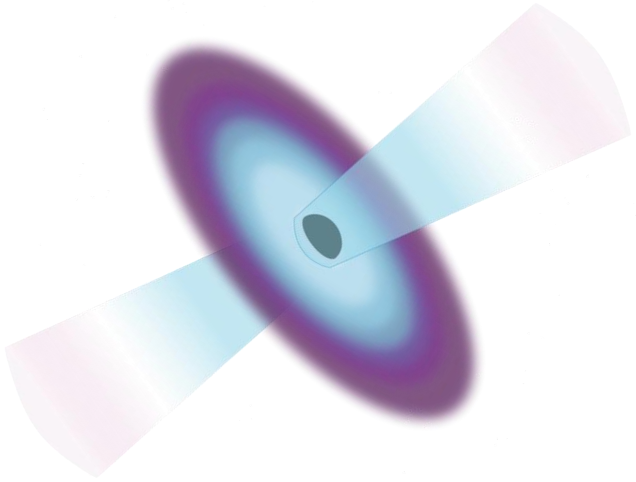


Truncation in Astronomy: Quasar Detection

Q: *Were older quasars brighter?*

Quasars are hard to detect! [Schafer, 2007]





Density Estimation From Truncated Samples

Exponential Family

$$\{\mathcal{E}_{h,t}(\theta) \propto h(x) \cdot e^{-\theta^\top t(x)} \;:\; \theta \in \mathbb{R}^m\}$$

Target Parameter

$$\boldsymbol{\theta}^\star \in \mathbb{R}^m$$

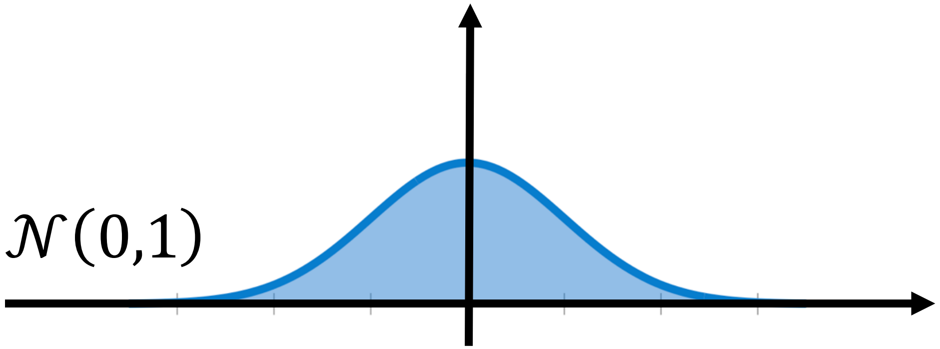
Density Estimation From Truncated Samples

Exponential Family

$$\{\mathcal{E}_{h,t}(\theta) \propto h(x) \cdot e^{-\theta^\top t(x)} \;:\; \theta \in \mathbb{R}^m\}$$

Target Parameter

$$\boldsymbol{\theta}^\star \in \mathbb{R}^m$$



Density Estimation From Truncated Samples

Density Estimation From Truncated Samples

Survival set: $S^\star \subseteq \mathbb{R}^d$

Density Estimation From Truncated Samples

Survival set: $S^\star \subseteq \mathbb{R}^d$

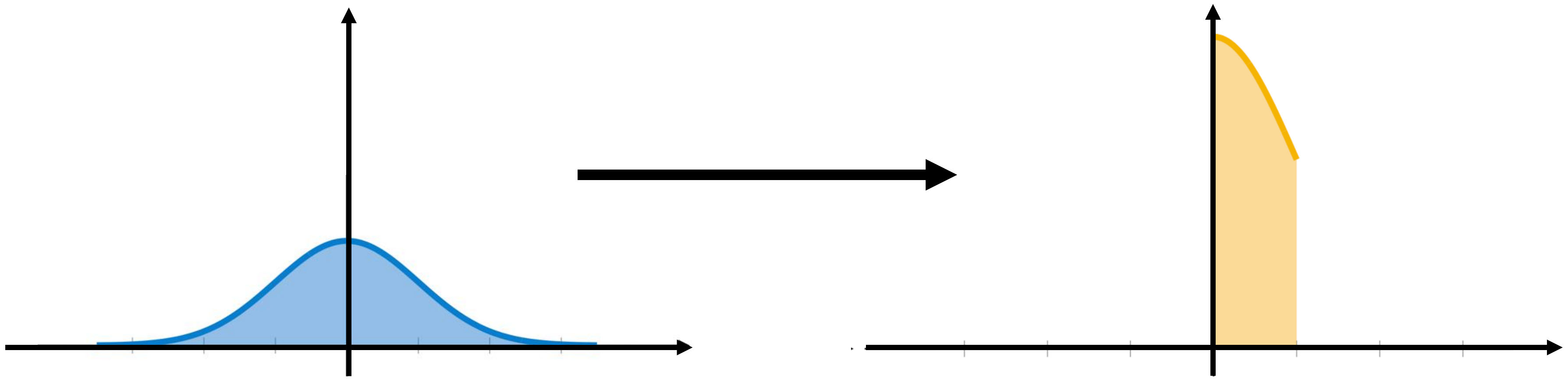
The mass of $S^\star$ is at least $\mathcal{E}(S^\star; \theta^\star) = \alpha$ where α is at least 1%

Density Estimation From Truncated Samples

Survival set: $S^\star \subseteq \mathbb{R}^d$
The mass of $S^\star$ is at least $\mathcal{E}(S^\star;\theta^\star) = \alpha$ where α is at least 1%

Samples observed from the truncated distribution

$$\mathcal{E}(\theta^\star, S^\star; x) = \frac{\mathbf{1}_{S^\star}(x)}{\alpha} \cdot \mathcal{E}(\theta^\star; x)$$



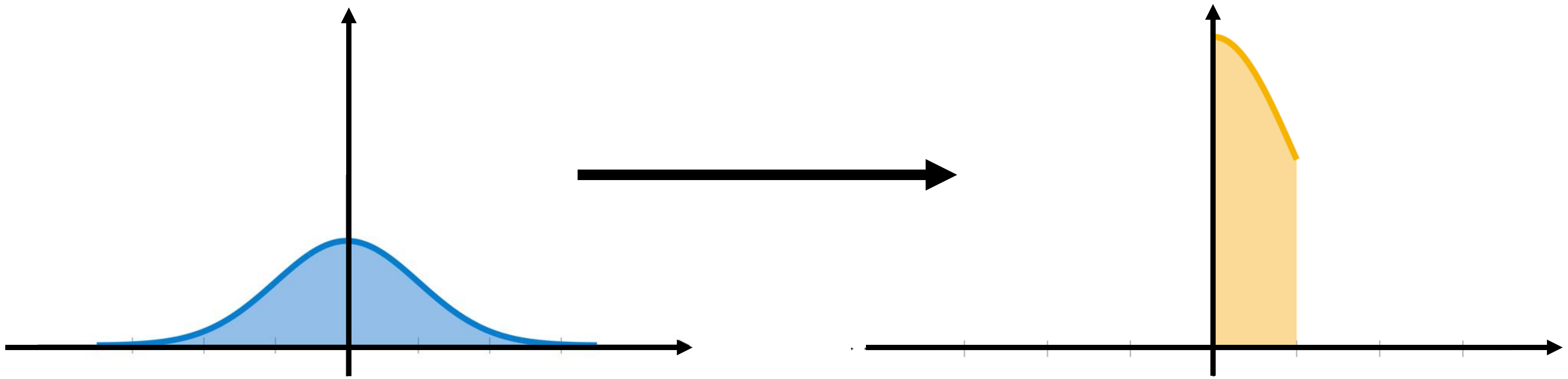
Density Estimation From Truncated Samples

Survival set: $S^\star \subseteq \mathbb{R}^d$

The mass of $S^\star$ is at least $\mathcal{E}(S^\star; \theta^\star) = \alpha$ where α is at least 1%

Samples observed from the truncated distribution

$$\mathcal{E}(\theta^\star, S^\star; x) = \frac{\mathbf{1}_{S^\star}(x)}{\alpha} \cdot \mathcal{E}(\theta^\star; x)$$



Goal: Given $x \sim \mathcal{E}(\boldsymbol{\theta}^\star, S^\star)$, find $\boldsymbol{\theta}$ such that $d_{TV}\big(\mathcal{E}(\boldsymbol{\theta}), \mathcal{E}(\boldsymbol{\theta}^\star)\big) \leq \varepsilon$

Prior Works

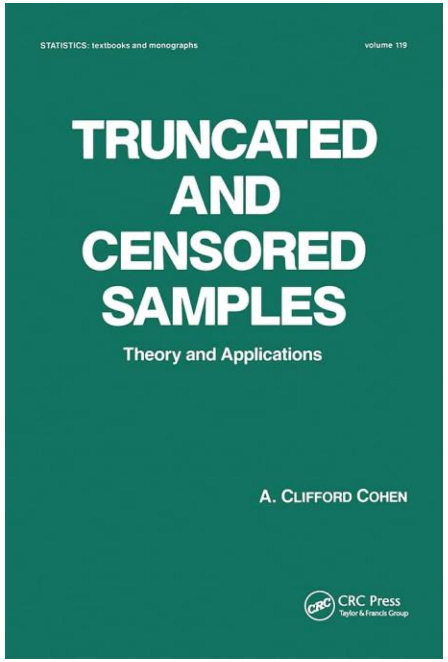
Long history in the Statistics and Economics Literature

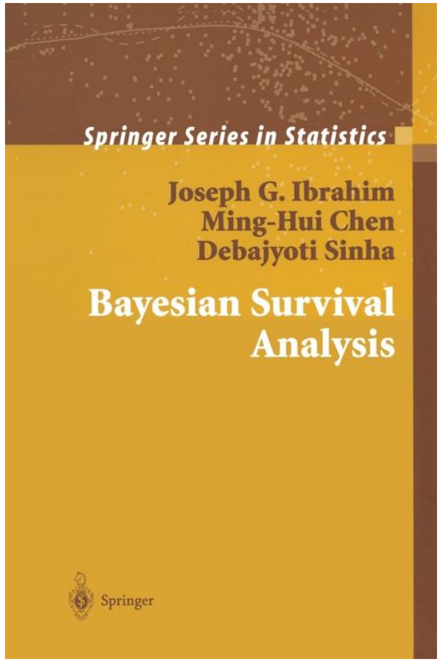
- Galton (1897), Pearson (1902), Pearson & Lee (1908), Fisher (1931), …, Maddala (1983)…
- *Computationally inefficient methods*

Prior Works

Long history in the Statistics and Economics Literature

- Galton (1897), Pearson (1902), Pearson & Lee (1908), Fisher (1931), …, Maddala (1983)…
- *Computationally inefficient methods*





Prior Works

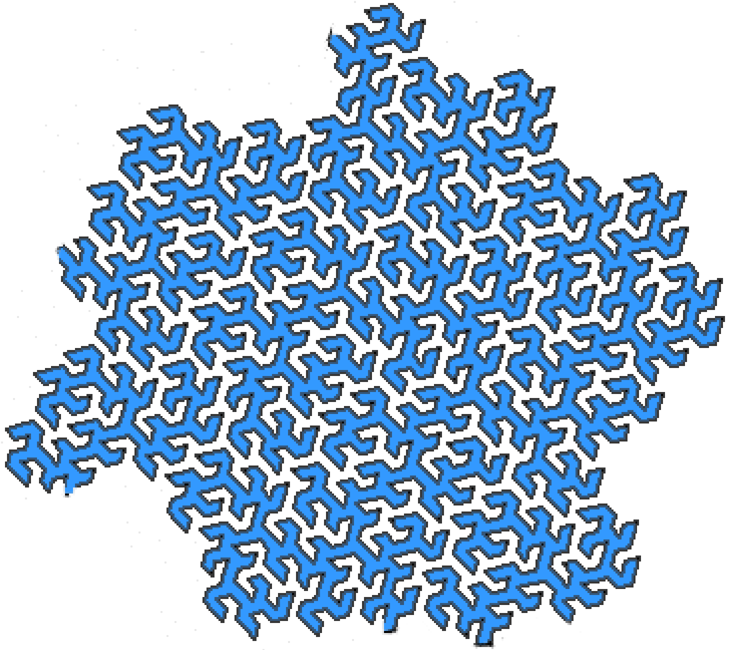
$O(d^2/\varepsilon^2)$ algorithms with "known" S^*

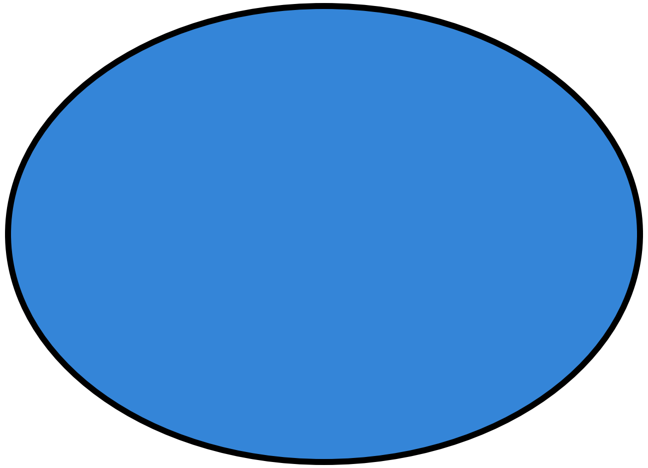
- Gaussians — Daskalakis, Gouleakis, Tzamos, and Zampetakis (FOCS 2018)
- Certain exponential families — Lee, Wibisono, Zampetakis (NeurIPS 2023)

Prior Works

Prior Works

Gaussian Surface Area $\Gamma(S)$: *Surface area with respect to the Gaussian measure*



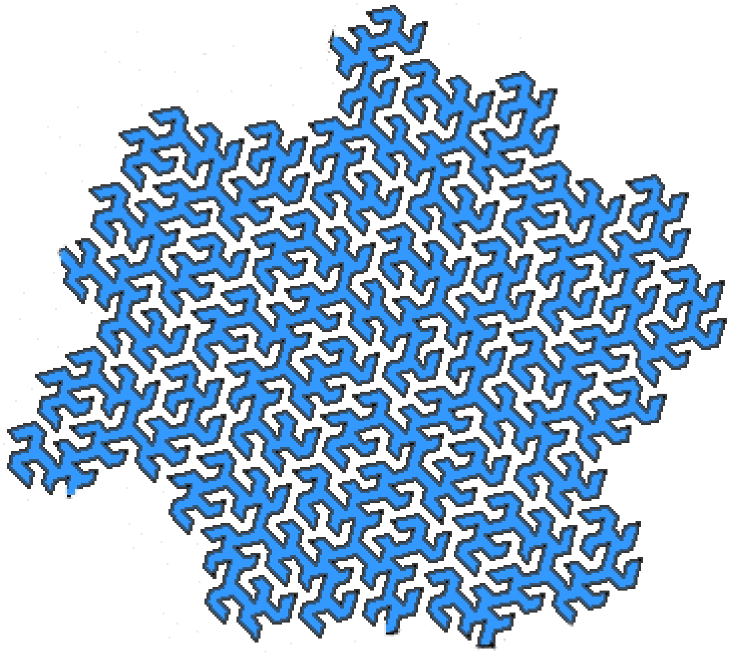


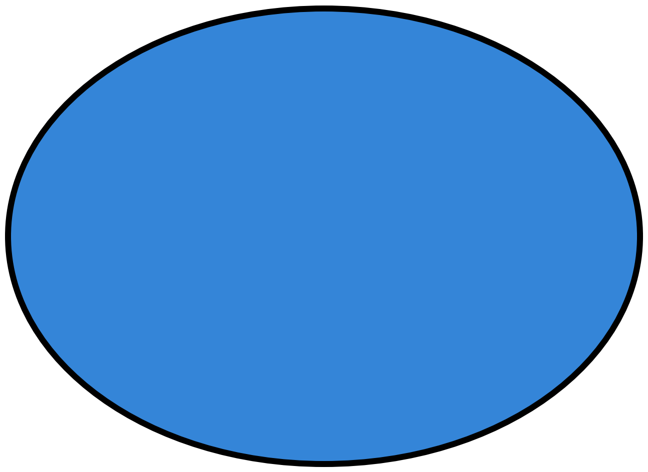
Prior Works

$d^{\text{poly}(\Gamma(S^\star)/\varepsilon)}$ algorithms with unknown S^*

- *Diagonal* gaussians — Kontonis, Tzamos, and Zampetakis (FOCS 2019)

Gaussian Surface Area $\Gamma(S)$: *Surface area with respect to the Gaussian measure*



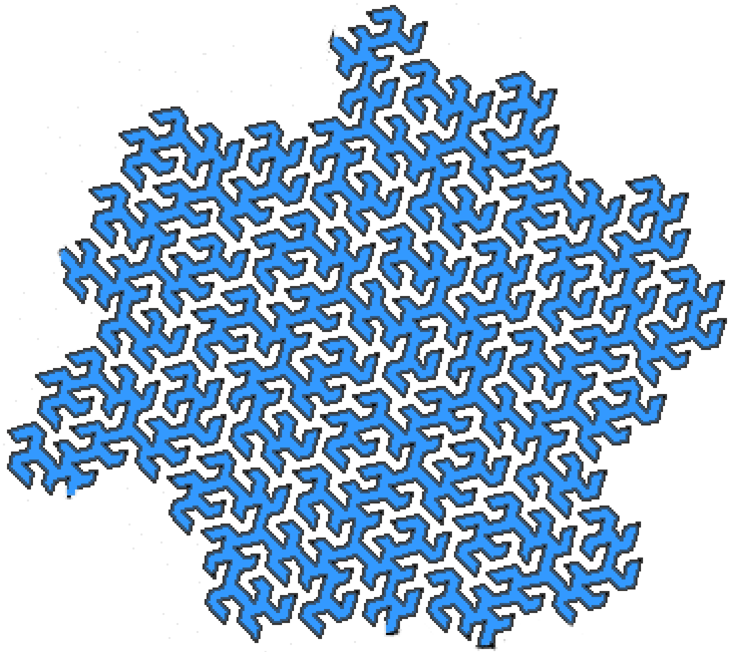


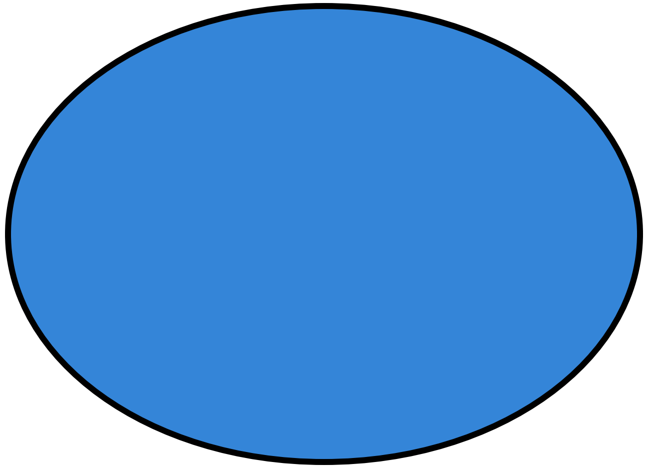
Prior Works

$d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ algorithms with unknown S^*

- *Diagonal* gaussians — Kontonis, Tzamos, and Zampetakis (FOCS 2019)
- SQ lower bound $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ — Diakonikolas, Kane, Pittas, Zarifis (COLT 2024)

Gaussian Surface Area $\Gamma(S)$: *Surface area with respect to the Gaussian measure*





Prior Works

$d^{\text{poly}(\Gamma(S^\star)/\varepsilon)}$ algorithms with unknown S^*

- *Diagonal* gaussians — Kontonis, Tzamos, and Zampetakis (FOCS 2019)
- SQ lower bound $d^{\text{poly}(\Gamma(S^\star)/\varepsilon)}$ — Diakonikolas, Kane, Pittas, Zarifis (COLT 2024)

Q1: Estimating *general* Gaussians with unknown S^*?

Prior Works

$d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ algorithms with unknown S^*

- *Diagonal* gaussians — Kontonis, Tzamos, and Zampetakis (FOCS 2019)
- SQ lower bound $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ — Diakonikolas, Kane, Pittas, Zarifis (COLT 2024)

Q1: Estimating *general* Gaussians with unknown S^*?

Q2: $\mathrm{poly}(d/\varepsilon)$ time estimation for halfspaces?

Results for Gaussians

Informal Theorem 1: In $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ time, we can find (μ, Σ), s.t., with 99% probability

$$d_{\mathrm{TV}}\left(\mathcal{N}(\mu,\Sigma),\ \mathcal{N}(\mu^\star,\Sigma^\star)\right) \le \varepsilon\ .$$

Results for Gaussians

Informal Theorem 1: In $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ time, we can find (μ, Σ), s.t., with 99% probability

$$\mathrm{d_{TV}}\big(\mathcal{N}(\mu,\Sigma),\ \mathcal{N}(\mu^\star,\Sigma^\star)\big) \le \varepsilon\ .$$

Informal Theorem 2: Let S^* be a halfspace or an axis-aligned box

In $\mathrm{poly}(d/\varepsilon)$ time, we can find (μ, Σ), s.t., with 99% probability

$$\mathrm{d_{TV}}\big(\mathcal{N}(\mu,\Sigma),\ \mathcal{N}(\mu^\star,\Sigma^\star)\big) \le \varepsilon\ .$$

Results for Gaussians

Informal Theorem 1: In $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ time, we can find (μ, Σ), s.t., with 99% probability

$$\mathrm{d_{TV}}\big(\mathcal{N}(\mu,\Sigma),\ \mathcal{N}(\mu^\star,\Sigma^\star)\big) \le \varepsilon\,.$$

Informal Theorem 2: Let S^* be a halfspace or an axis-aligned box

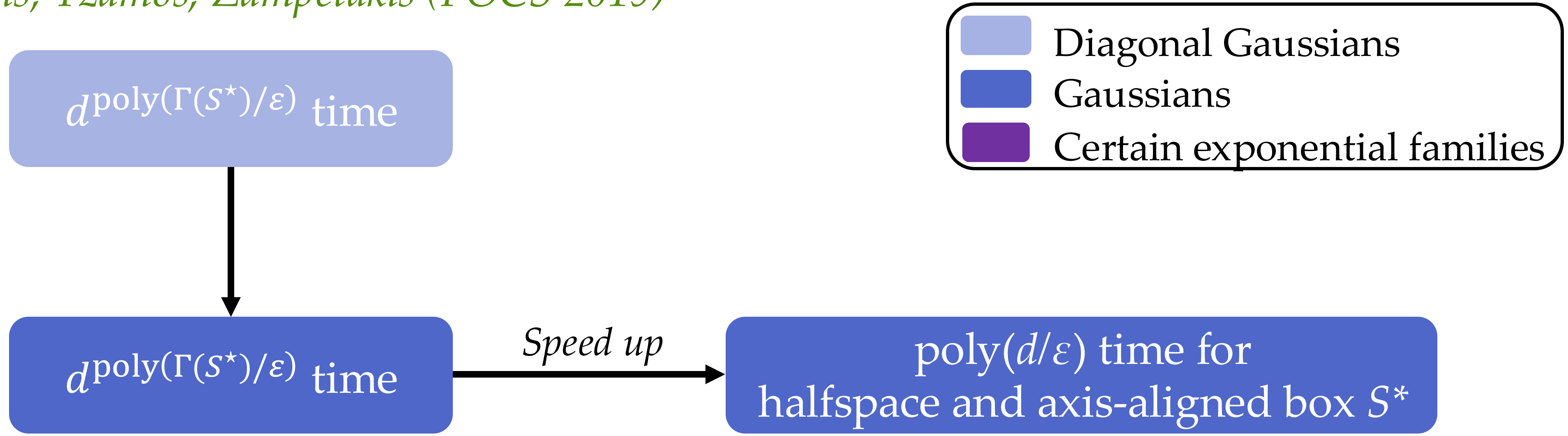
In $\mathrm{poly}(d/\varepsilon)$ time, we can find (μ, Σ), s.t., with 99% probability

$$\mathrm{d_{TV}}\big(\mathcal{N}(\mu,\Sigma),\ \mathcal{N}(\mu^\star,\Sigma^\star)\big) \le \varepsilon\,.$$

Both results extend to truncated linear regression with Gaussian covariates with unknown S^*

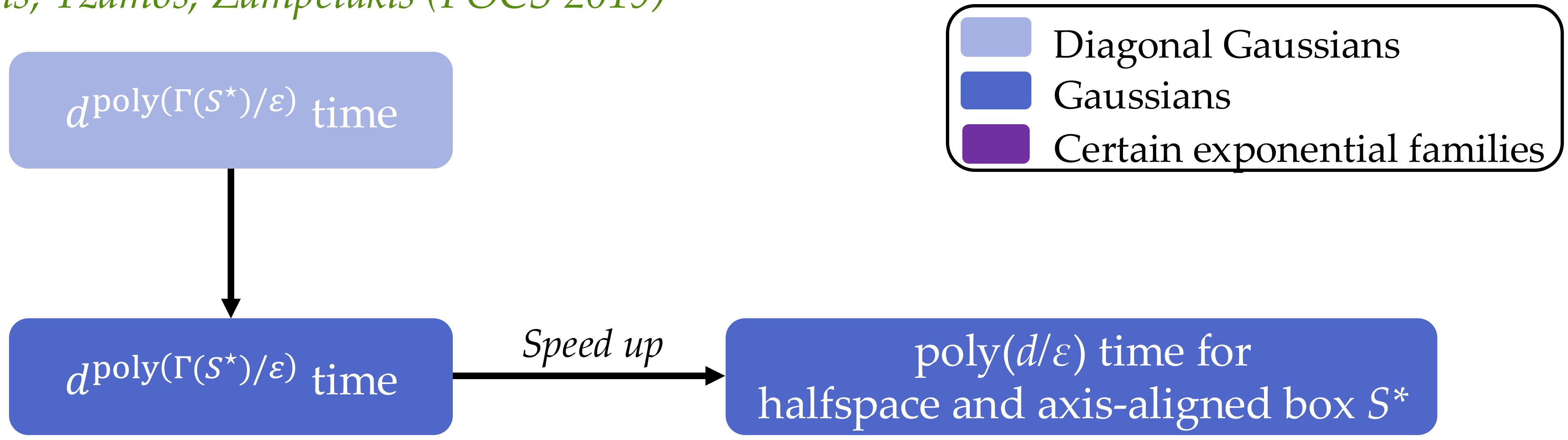
Results Beyond Gaussians

Kontonis, Tzamos, Zampetakis (FOCS 2019)



Results Beyond Gaussians

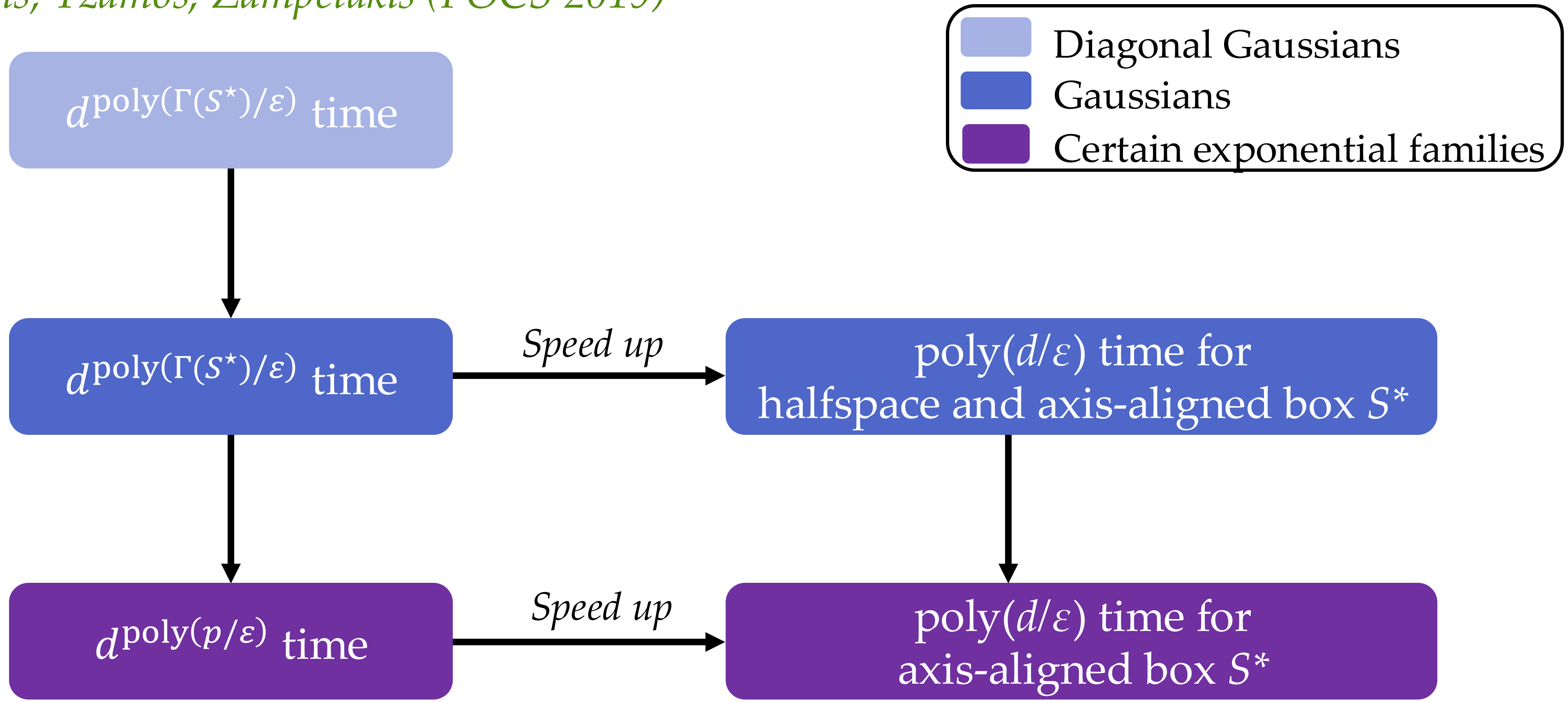
Kontonis, Tzamos, Zampetakis (FOCS 2019)



Our approach is <u>not</u> based on Hermite polynomials

Results Beyond Gaussians

Kontonis, Tzamos, Zampetakis (FOCS 2019)



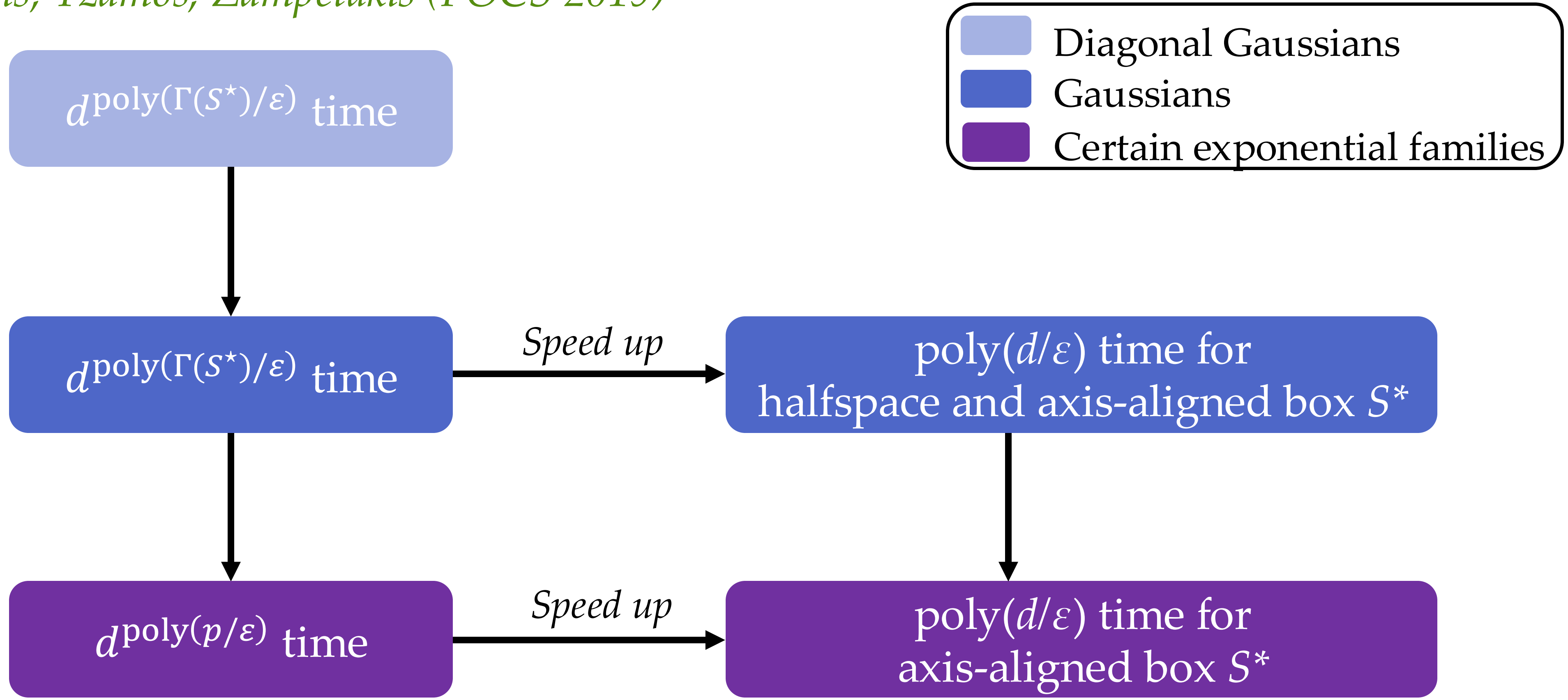
Our approach is <u>not</u> based on Hermite polynomials

Degree-p polynomials poly(ε)-approximate S^* with respect to $\mathcal{E}(\theta^\star)$ in L_2-norm

Results Beyond Gaussians

Kontonis, Tzamos, Zampetakis (FOCS 2019)

Our approach is not based on Hermite polynomials



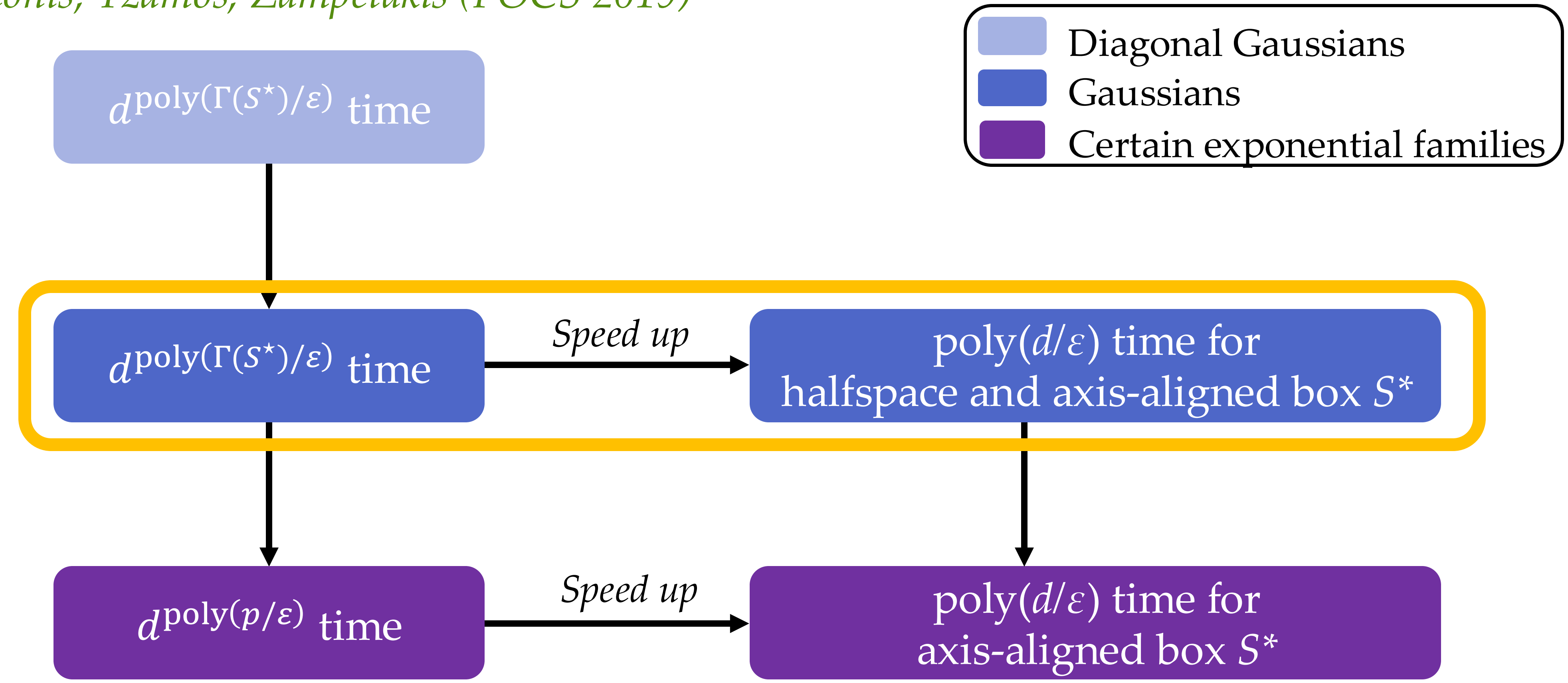
Degree-p polynomials poly(ε)-approximate S^* with respect to $\mathcal{E}(\theta^\star)$ in L_2-norm

Open Question: *poly(d/ε) time algorithm for halfspaces?*

Results Beyond Gaussians

Kontonis, Tzamos, Zampetakis (FOCS 2019)

Our approach is <u>not</u> based on Hermite polynomials



Degree-p polynomials poly(ε)-approximate S^* with respect to $\mathcal{E}(\theta^\star)$ in L_2-norm

Open Question: *poly(d/ε) time algorithm for halfspaces?*

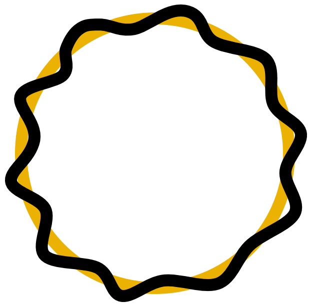
Proof Outline

$LL \equiv$ Log-likelihood

Proof Outline

$LL \equiv$ Log-likelihood

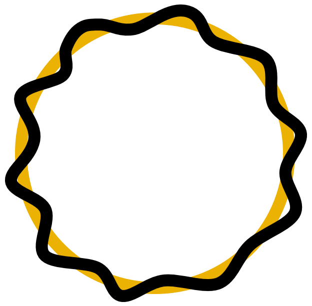
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$



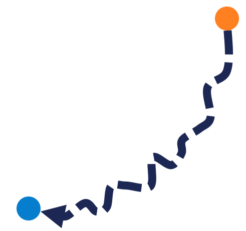
Proof Outline

$LL \equiv$ Log-likelihood

(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$



(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

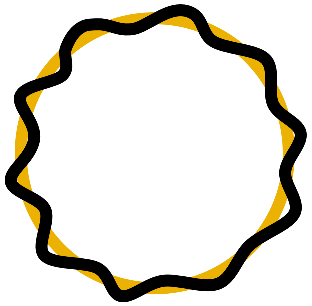


Proof Outline

$LL \equiv$ Log-likelihood

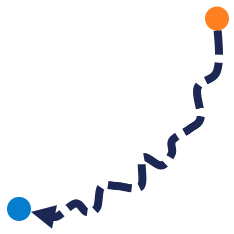
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^{\star}$

Time: $d^{\mathrm{poly}(\Gamma(S^{\star})/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$



(2) Learn $\theta^{\star}$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$

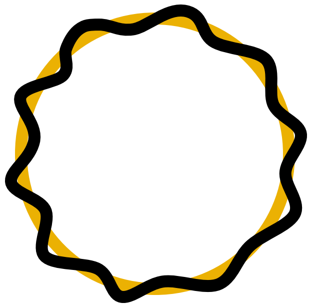


Proof Outline

$LL \equiv$ Log-likelihood

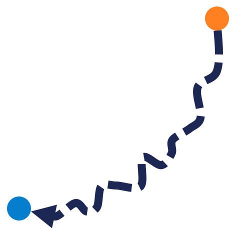
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$



(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



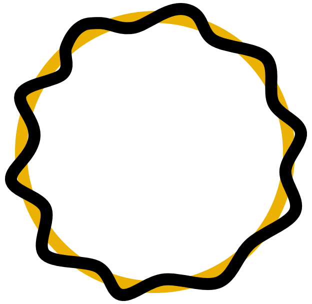
Challenge: Only have samples inside $S^\star$

Proof Outline

$LL \equiv$ Log-likelihood

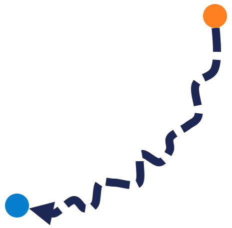
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^{\star}$

Time: $d^{\mathrm{poly}(\Gamma(S^{\star})/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

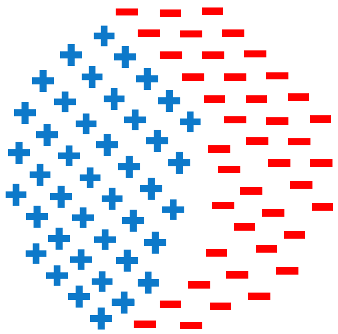


(2) Learn $\theta^{\star}$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^{*}

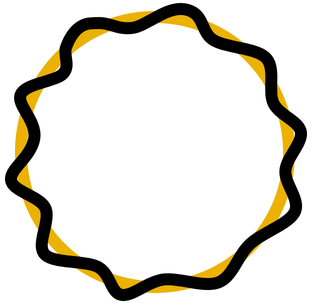


Proof Outline

$LL \equiv$ Log-likelihood

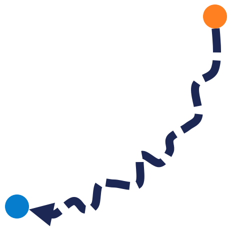
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^{\star}$

Time: $d^{\mathrm{poly}(\Gamma(S^{\star})/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

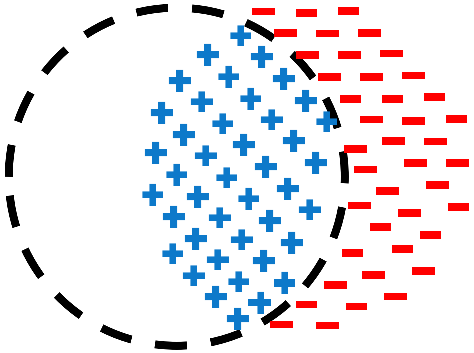


(2) Learn $\theta^{\star}$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^{*}

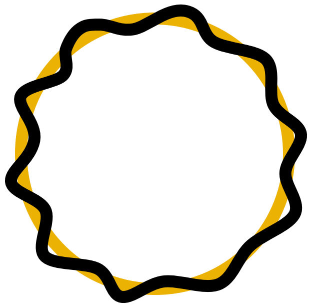


Proof Outline

$LL \equiv$ Log-likelihood

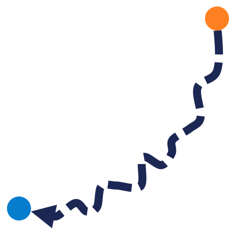
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

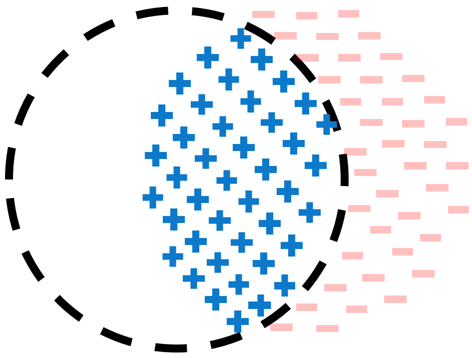


(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^*

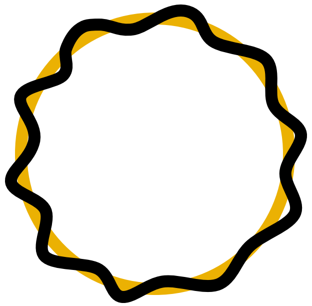


Proof Outline

$LL \equiv$ Log-likelihood

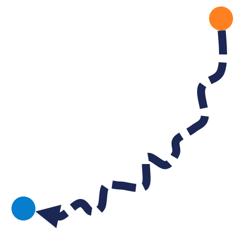
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

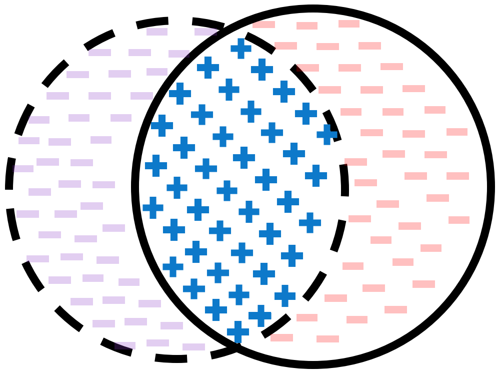


(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^*

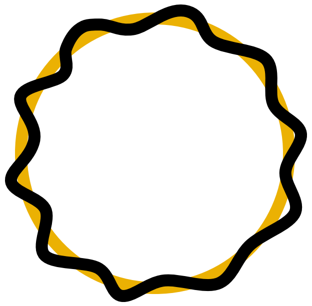


Proof Outline

$LL \equiv$ Log-likelihood

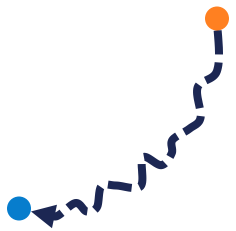
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

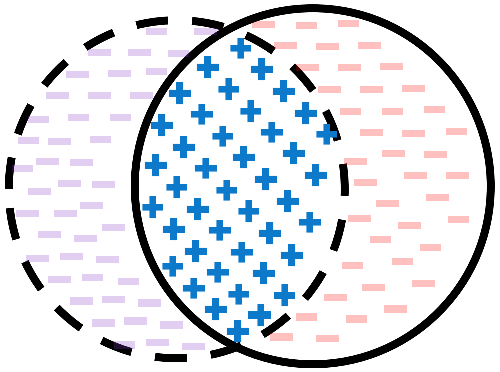


(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^*

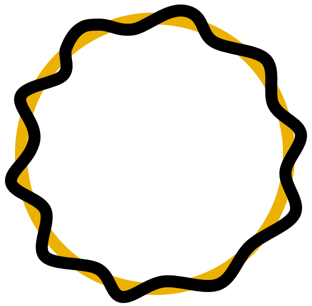


Almost *no classes* can be learned from just positive samples (Natarajan, STOC, 1987)

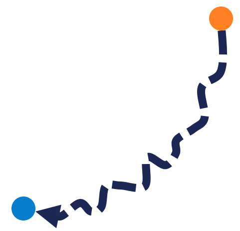
Proof Outline

$LL \equiv$ Log-likelihood

(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$



(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

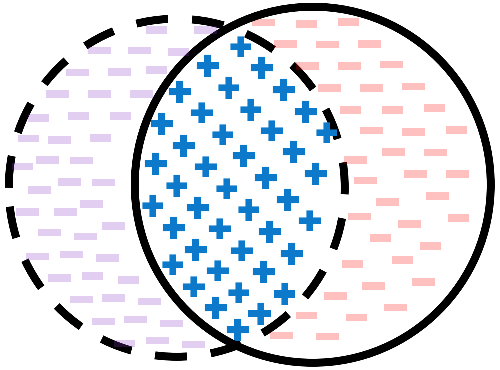


Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

Time: $\mathrm{poly}(d/\varepsilon)$

Challenge: Only have samples inside S^*

Challenge: LL with $S^* \rightarrow S$ is ∞ everywhere



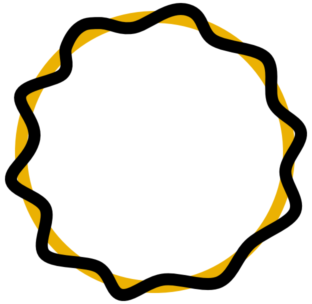
Almost *no classes* can be learned from just positive samples (Natarajan, STOC, 1987)

Proof Outline

$LL \equiv$ Log-likelihood

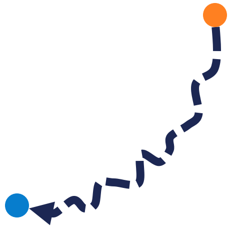
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

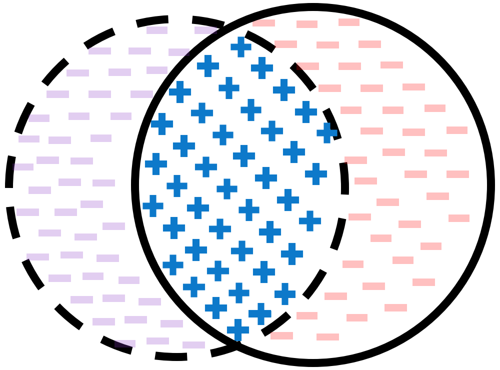


(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^*



Almost *no classes* can be learned from just positive samples (Natarajan, STOC, 1987)

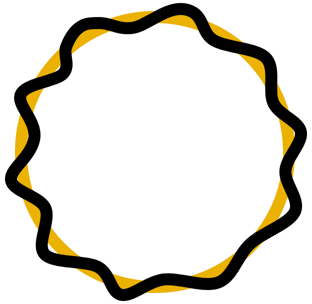
Challenge: LL with $S^* \rightarrow S$ is ∞ everywhere

Proof Outline

LL ≡ Log-likelihood

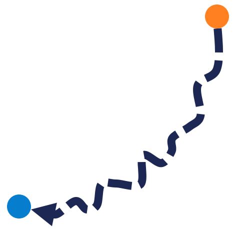
(1) Learn $\boldsymbol{S} \approx \boldsymbol{S}^\star$

Time: $d^{\mathrm{poly}(\Gamma(S^\star)/\varepsilon)}$ or $\mathrm{poly}(d/\varepsilon)$

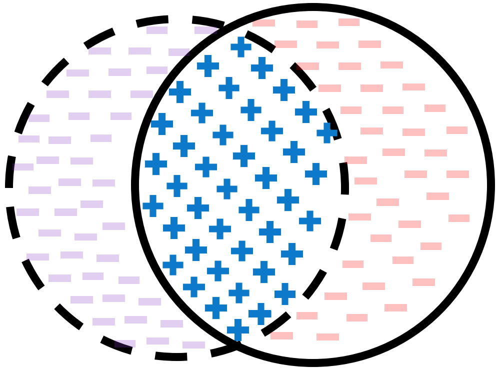


(2) Learn $\theta^\star$ with access to $\boldsymbol{S}$

Time: $\mathrm{poly}(d/\varepsilon)$



Challenge: Only have samples inside S^*

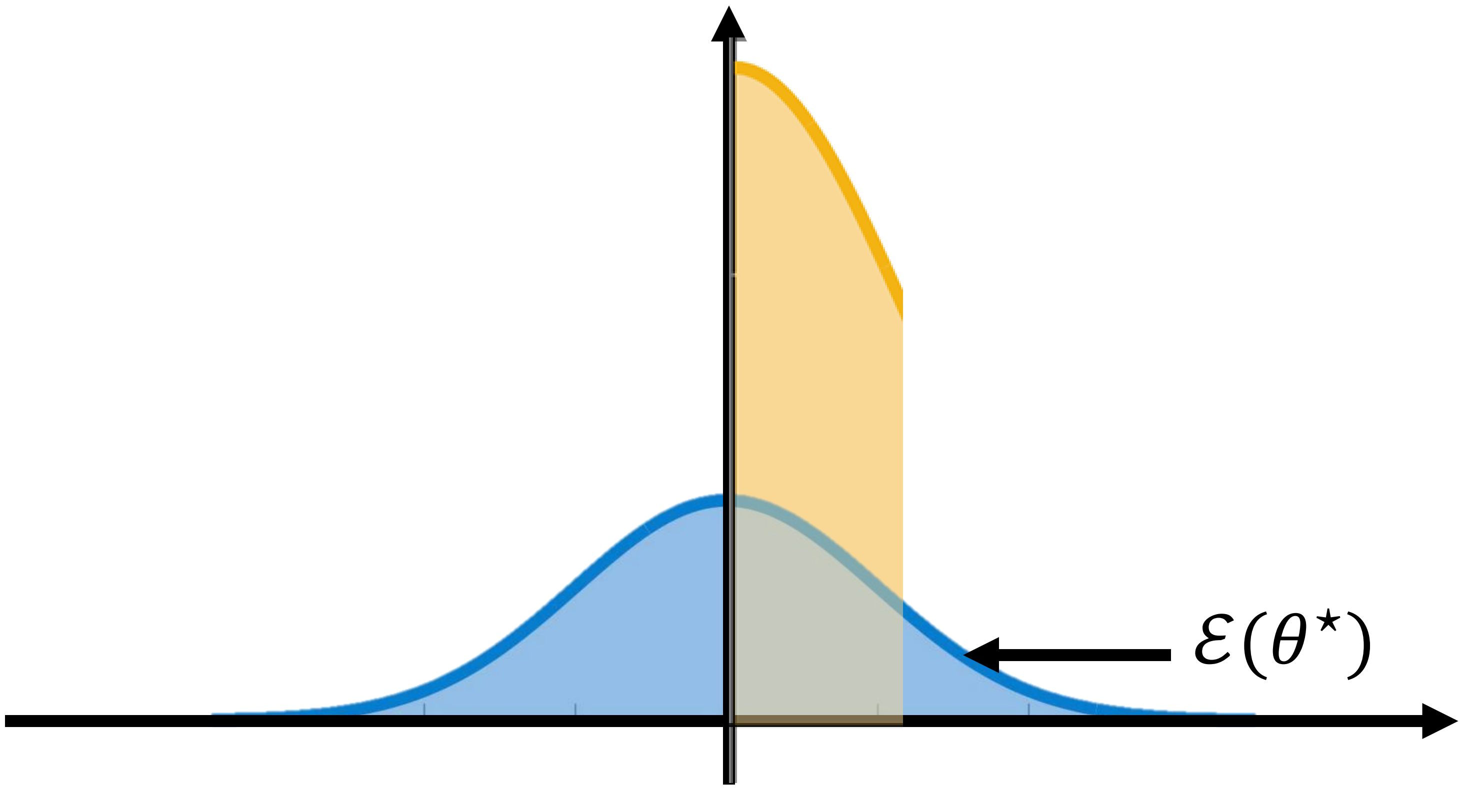


Almost *no classes* can be learned from just positive samples (Natarajan, STOC, 1987)

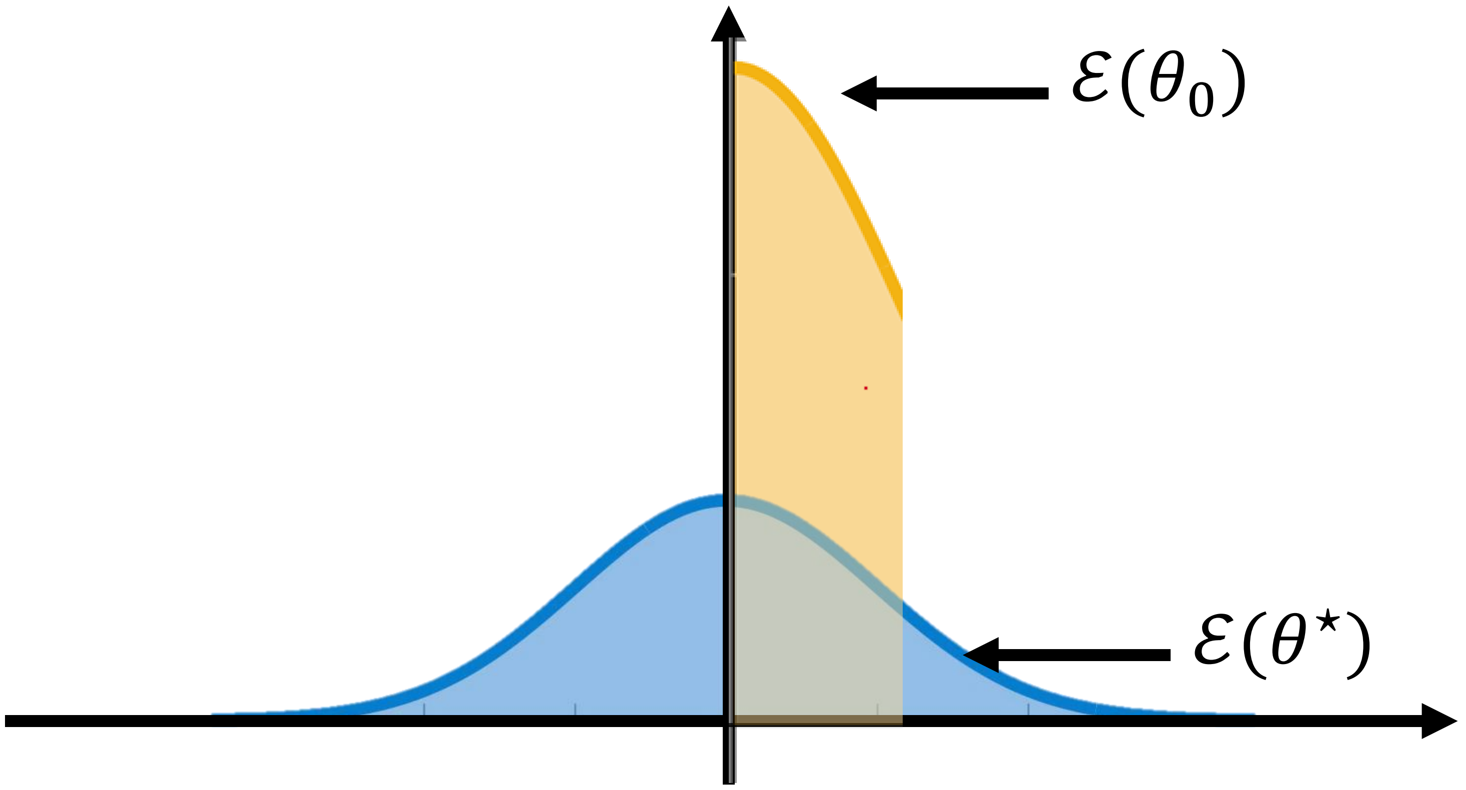
Challenge: LL with $S^* \to S$ is ∞ everywhere

A new optimization program:
Perturbed log-likelihood (PLL)

Obtaining Unlabeled Samples



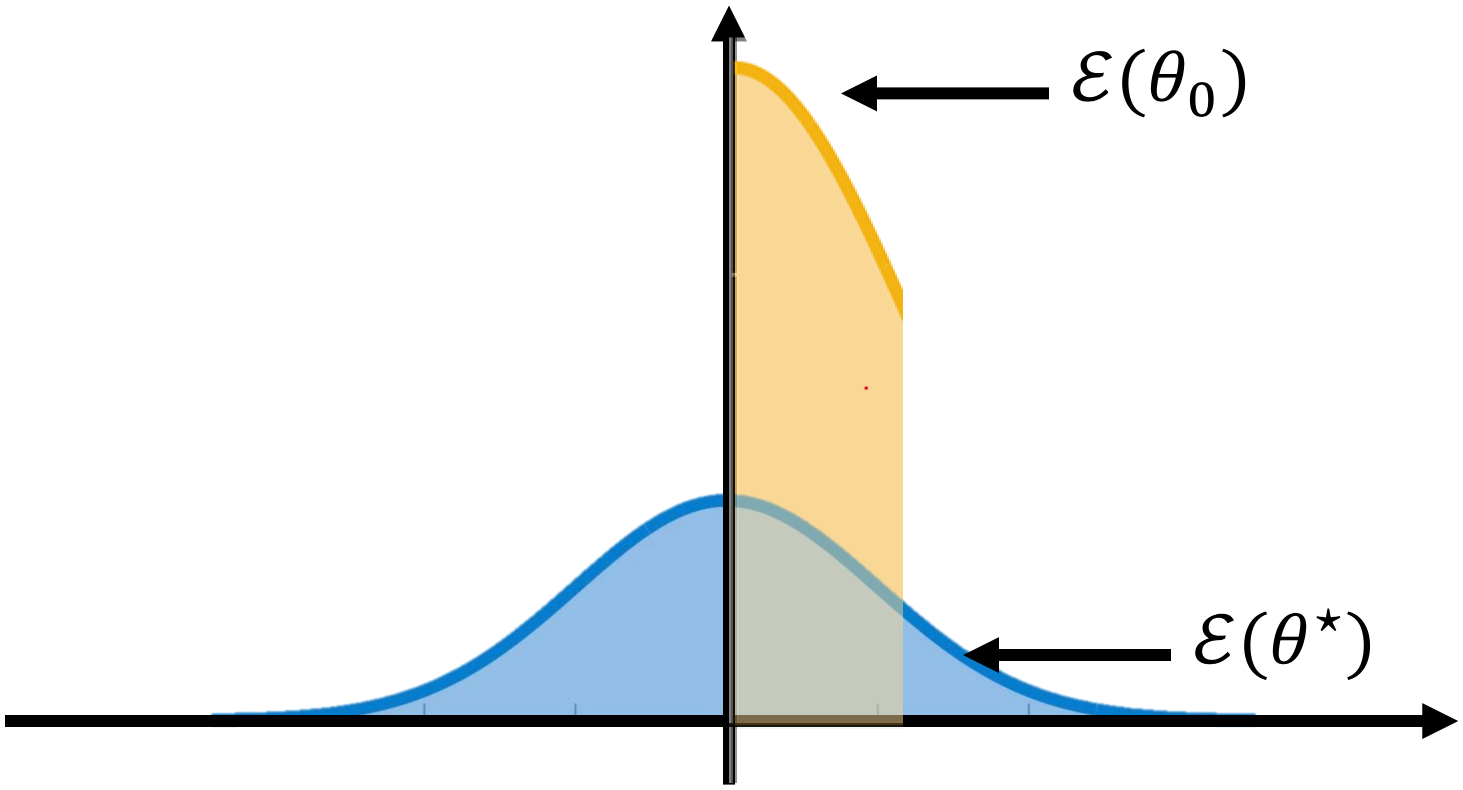
Obtaining Unlabeled Samples



Obtaining Unlabeled Samples

$\mathcal{E}(\theta^\star)$ and $\mathcal{E}(\theta_0)$ are C-close for $C = \mathrm{poly}(1/\alpha)$:

$$\text{for all sets } T \subseteq \mathbb{R}^d, \qquad e^{-C} \cdot \mathcal{E}(\theta_0; T) \le \mathcal{E}(\theta^\star; T) \le e^{C} \cdot \mathcal{E}(\theta_0; T)^{1/C}$$



Robustly Learning from Positive-Unlabeled Samples

$\mathcal{E}(\theta^\star)$ and $\mathcal{E}(\theta_0)$ are C-close for $C = \text{poly}(1/\alpha)$:

$$\text{for all sets } T \subseteq \mathbb{R}^d, \qquad e^{-C} \cdot \mathcal{E}(\theta_0; T) \leq \mathcal{E}(\theta^\star; T) \leq e^{C} \cdot \mathcal{E}(\theta_0; T)^{1/\mathrm{C}}$$

Robustly Learning from Positive-Unlabeled Samples

$\mathcal{E}(\theta^\star)$ and $\mathcal{E}(\theta_0)$ are C-close for $C = \text{poly}(1/\alpha)$:

$$\text{for all sets } T \subseteq \mathbb{R}^d, \qquad e^{-C} \cdot \mathcal{E}(\theta_0; T) \leq \mathcal{E}(\theta^\star; T) \leq e^{C} \cdot \mathcal{E}(\theta_0; T)^{1/C}$$

Informal Theorem: $\mathcal{H}$ can PAC-learned from

- n iid positive samples and
- n iid unlabeled samples from C-close to $\mathcal{D}_X$

for

$$n = \tilde{O}\left(\frac{1}{\varepsilon^{2+2C}} \cdot (\text{VC}(\mathcal{H}) + \log \delta^{-1})\right)$$

Robustly Learning from Positive-Unlabeled Samples

$\mathcal{E}(\theta^\star)$ and $\mathcal{E}(\theta_0)$ are C-close for $C = \mathrm{poly}(1/\alpha)$:

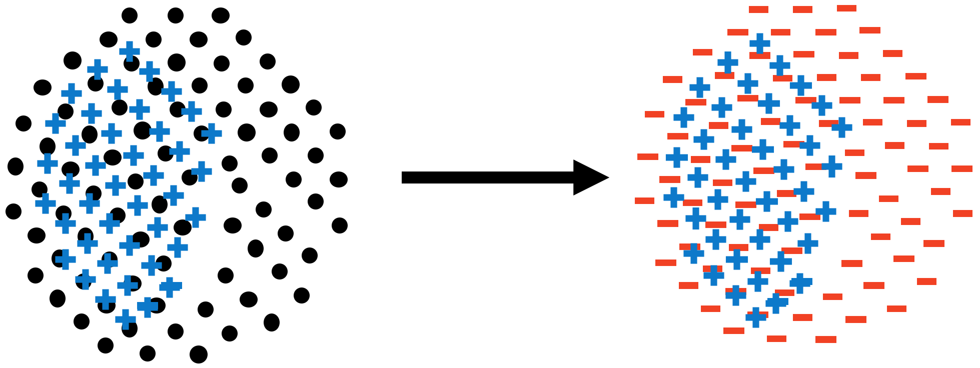
$$\text{for all sets } T \subseteq \mathbb{R}^d, \qquad e^{-C} \cdot \mathcal{E}(\theta_0; T) \le \mathcal{E}(\theta^\star; T) \le e^{C} \cdot \mathcal{E}(\theta_0; T)^{1/C}$$

Informal Theorem: $\mathcal{H}$ can PAC-learned from

- n iid positive samples and
- n iid unlabeled samples from C-close to $\mathcal{D}_X$

for

$$n = \tilde{O}\left(\frac{1}{\varepsilon^{2+2C}} \cdot \left(\mathrm{VC}(\mathcal{H}) + \log \delta^{-1}\right)\right)$$



Efficiently Learning S^*: General sets

Algorithm: L_1-Regression *(Kalai, Klivans, Mansour, Servedio, FOCS 2005)*

L_1-Regression whenever $\mathbf{1}_{S^\star}(x)$ can be approximated by polynomials w.r.t. $\mathcal{E}(\theta^\star)$

Efficiently Learning S^*: General sets

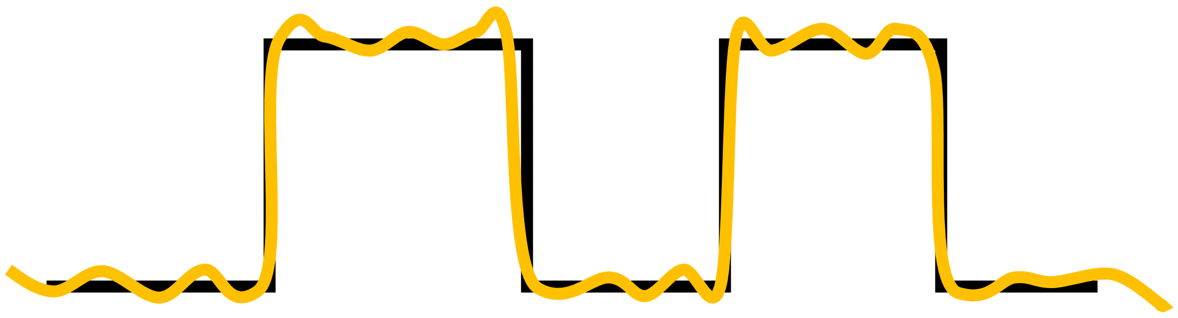
Algorithm: L_1-Regression *(Kalai, Klivans, Mansour, Servedio, FOCS 2005)*

L_1-Regression whenever $\mathbf{1}_{S^\star}(x)$ can be approximated by polynomials w.r.t. $\mathcal{E}(\theta^\star)$

Efficiently Learning S^*: General sets

Algorithm: L_1-Regression *(Kalai, Klivans, Mansour, Servedio, FOCS 2005)*

L_1-Regression whenever $\mathbf{1}_{S^\star}(x)$ can be approximated by polynomials w.r.t. $\mathcal{E}(\theta^\star)$

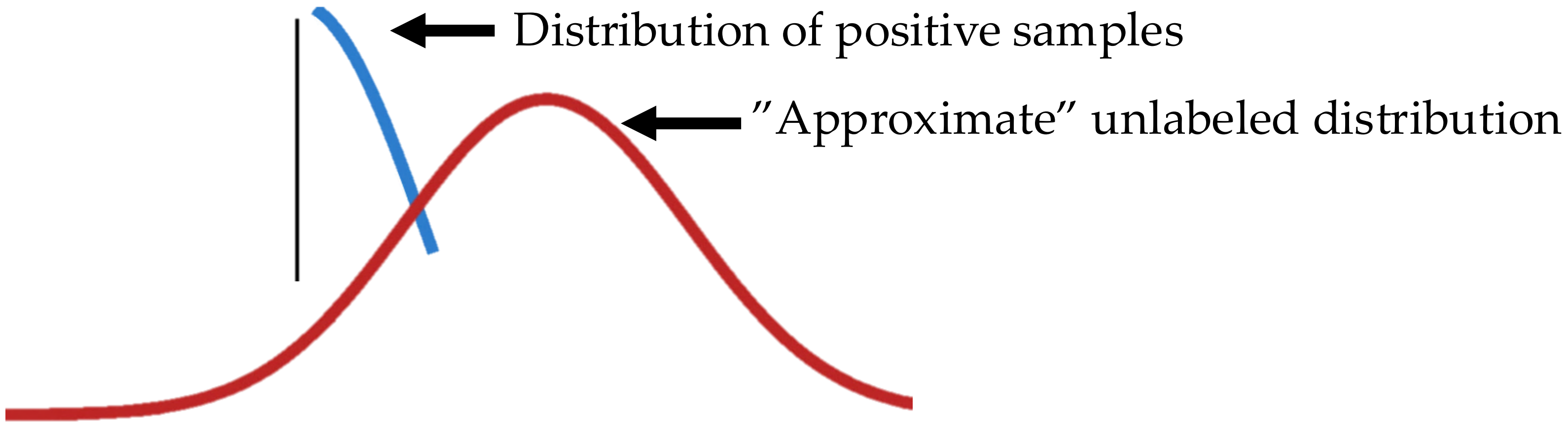


Efficiently Learning S^*: General Sets

Most approximability results are known w.r.t. log-concave distributions

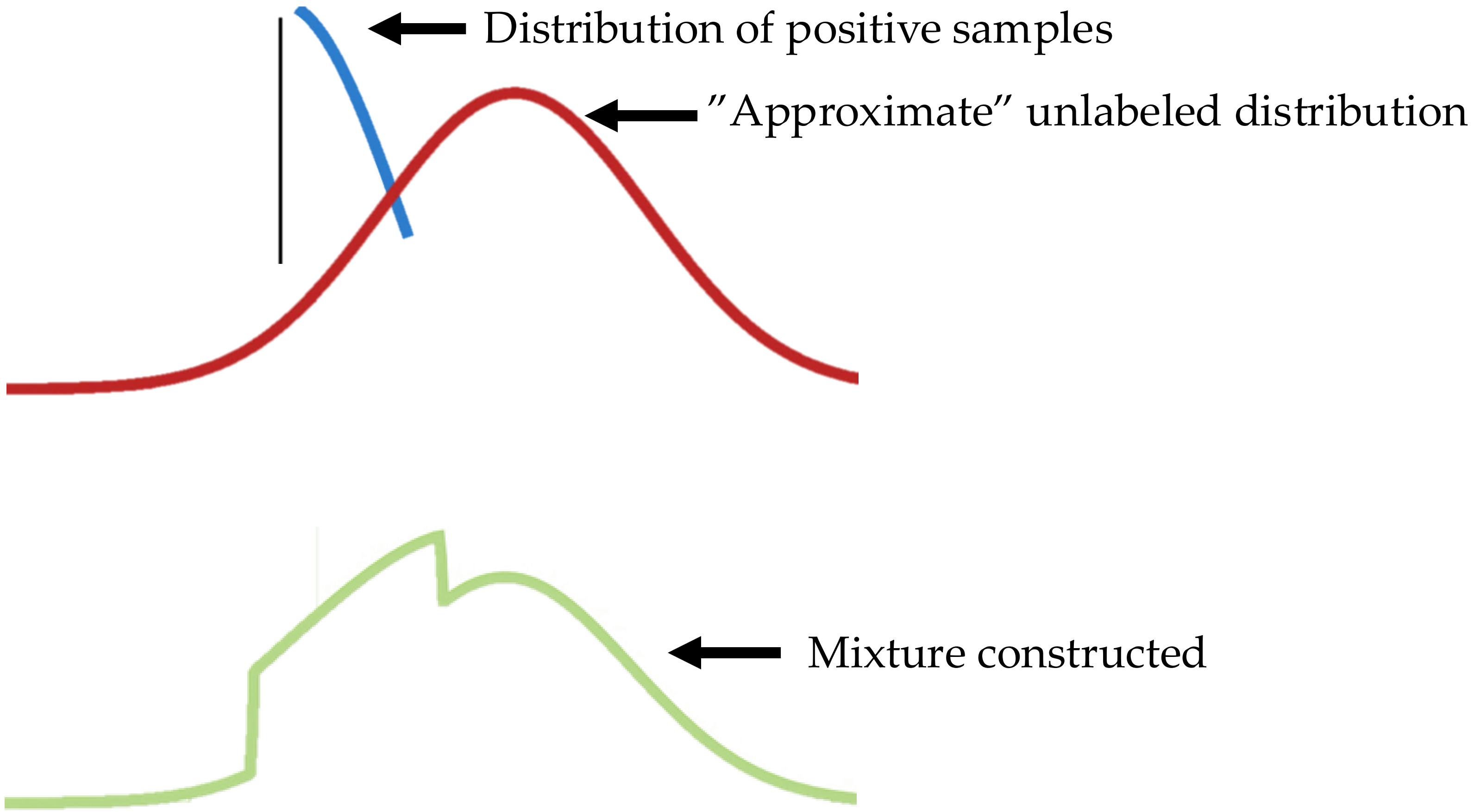
Efficiently Learning S^*: General Sets

Most approximability results are known w.r.t. log-concave distributions



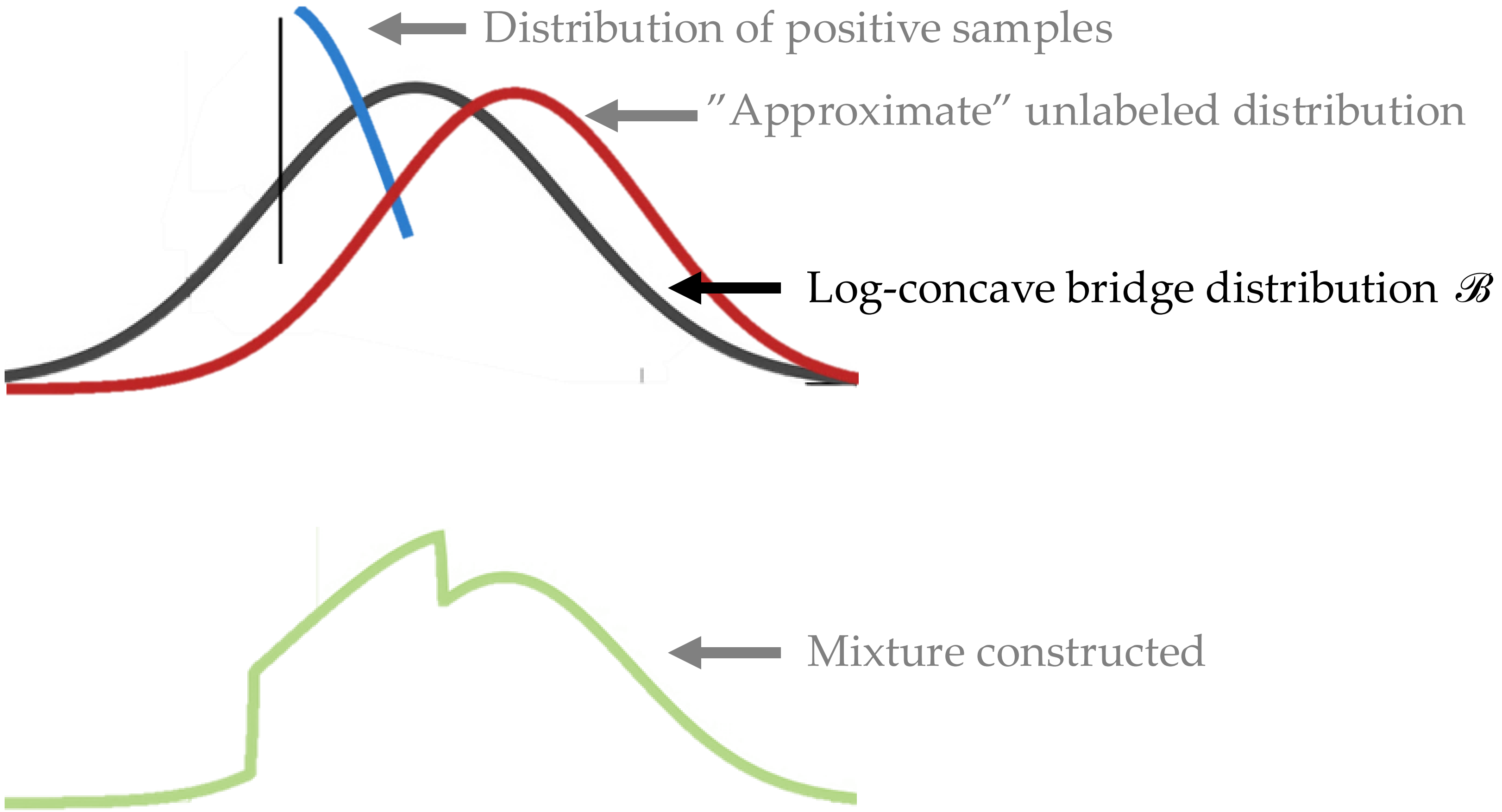
Efficiently Learning S^*: General Sets

Most approximability results are known w.r.t. log-concave distributions



Efficiently Learning S^*: General Sets

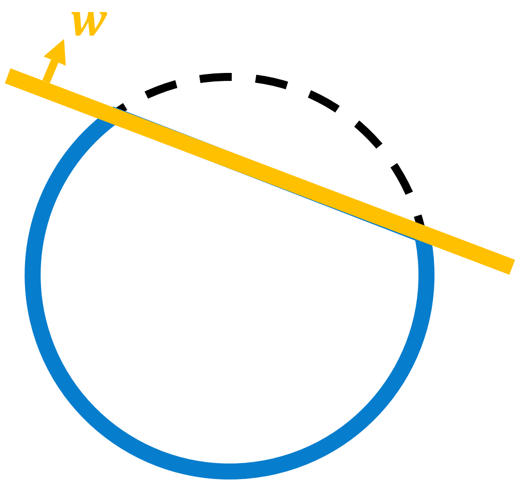
Most approximability results are known w.r.t. log-concave distributions



Efficiently Learning S^*: Specific Families

Halfspaces

Moment-Based Method to learn halfspaces w.r.t. unknown Gaussian

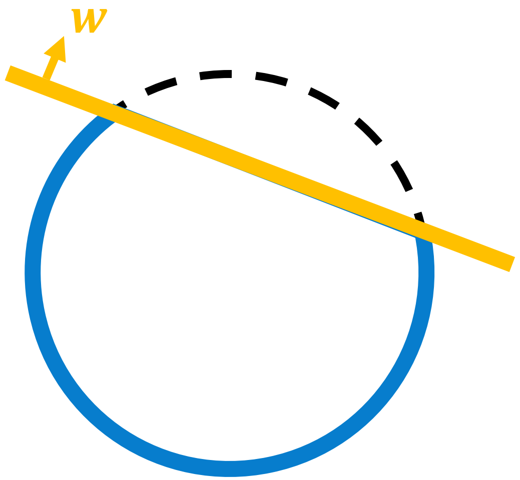


Efficiently Learning S^*: Specific Families

Halfspaces

Moment-Based Method to learn halfspaces w.r.t. unknown Gaussian

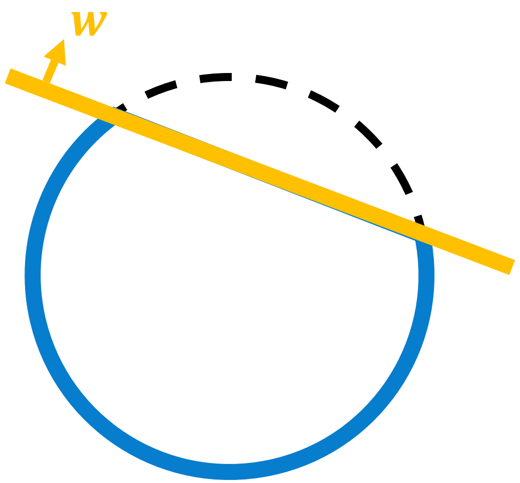
$$\mathbb{E}_x[x] \quad \propto \quad \mu^\star + \boldsymbol{w}$$



Efficiently Learning S^*: Specific Families

Halfspaces

Moment-Based Method to learn halfspaces w.r.t. unknown Gaussian



$$\mathbb{E}_x[x] \quad \propto \quad \mu^\star + \boldsymbol{w}$$

$$\mathrm{Cov}_x[x] \quad \propto \quad \Sigma^\star + \boldsymbol{w}\boldsymbol{w}^\top$$

Efficiently Learning S^*: Specific Families

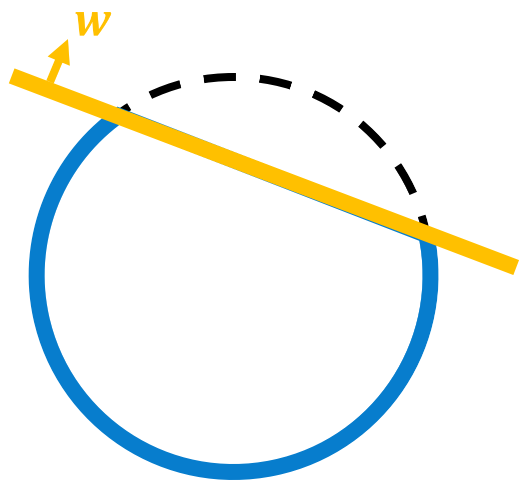
Halfspaces

Moment-Based Method to learn halfspaces w.r.t. unknown Gaussian

$$\mathbb{E}_x[x] \quad \propto \quad \mu^\star + \boldsymbol{w}$$

$$\mathrm{Cov}_x[x] \quad \propto \quad \Sigma^\star + \boldsymbol{w}\boldsymbol{w}^\top$$

$$\mathbb{E}_x\left[\left(x - \mu_{x,S^\star}\right)^{\otimes 3}\right] \quad \propto \quad \boldsymbol{w}^{\otimes 3}$$



Efficiently Learning S^*: Specific Families

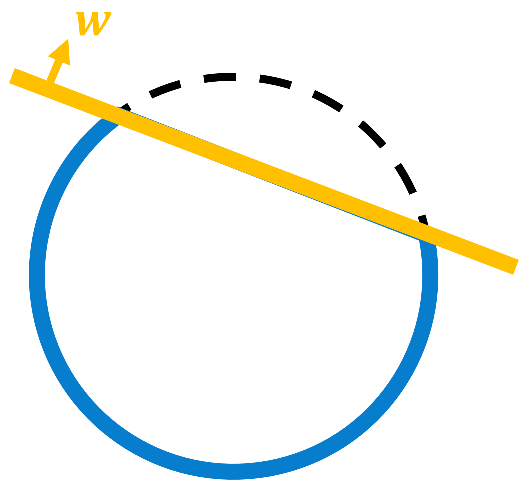
Halfspaces

Moment-Based Method to learn halfspaces w.r.t. unknown Gaussian

$$\mathbb{E}_x[x] \propto \mu^\star + \boldsymbol{w}$$

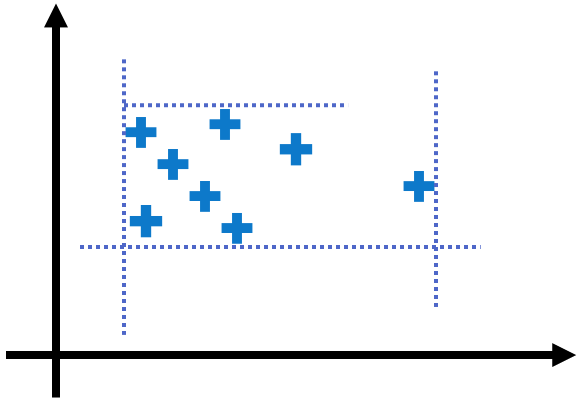
$$\mathrm{Cov}_x[x] \propto \Sigma^\star + \boldsymbol{w}\boldsymbol{w}^\top$$

$$\mathbb{E}_x\left[\left(x - \mu_{x,S^\star}\right)^{\otimes 3}\right] \propto \boldsymbol{w}^{\otimes 3}$$



Axis-Aligned Rectangle

Folklore Method to learning from only positive samples



Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and "imperfect" unlabeled samples to agnostic learning

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and "imperfect" unlabeled samples to agnostic learning

Open Problems:

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and "imperfect" unlabeled samples to agnostic learning

Open Problems:

- poly(d/ε) algorithm for other classes, e.g., intersections of (two) halfspaces

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and "imperfect" unlabeled samples to agnostic learning

Open Problems:

- poly(d/ε) algorithm for other classes, e.g., intersections of (two) halfspaces
- Truncated Linear regression with non-Gaussian features

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and "imperfect" unlabeled samples to agnostic learning

Open Problems:

- poly(d/ε) algorithm for other classes, e.g., intersections of (two) halfspaces
- Truncated Linear regression with non-Gaussian features
- Unknown truncation with weaker requirements than bounds on GSA

Summary and Open Problems

New Results in Truncated Statistics with Unknown Truncation

- First efficient algorithm for general Gaussians (and beyond)
- First efficient algorithm for truncated linear regression with Gaussian features
- First poly(d/ε) algorithm for restricted truncation sets (e.g., halfspaces)

A reduction from learning with positive and “imperfect” unlabeled samples to agnostic learning

Open Problems:

- poly(d/ε) algorithm for other classes, e.g., intersections of (two) halfspaces
- Truncated Linear regression with non-Gaussian features
- Unknown truncation with weaker requirements than bounds on GSA

Thank You!